

ABSTRAK

Pada mesin pencari *web*, jumlah dokumen relevan yang diperoleh dipengaruhi oleh jumlah kata kunci yang ada dalam *query*. *Query* yang dimasukkan oleh pengguna saat melakukan pencarian pada umumnya pendek, terdiri dari satu atau dua kata saja. Hal ini mengakibatkan dokumen relevan yang terpanggil pada *database* hanya sedikit karena sedikitnya informasi dalam *query*. Pada penelitian ini sistem temu kembali informasi pariwisata dibangun menggunakan *vector space model* dan *query expansion* menggunakan kamus Tesaurus Bahasa Indonesia. *Vector space model* merupakan model yang digunakan untuk menyimpan dokumen dalam bentuk matriks *term-document*. Sedangkan *query expansion* merupakan teknik untuk memperluas makna dari kata kunci agar memenuhi kebutuhan informasi yang dibutuhkan oleh pengguna. Informasi halaman *web* sebelumnya diambil menggunakan *web crawler* dan tersimpan di *database*. Pengukuran kinerja dilakukan dengan cara menghitung nilai *precision*, *recall* dan *f-measure* dari hasil pencarian dengan beberapa kata kunci yang berbeda. Dari hasil pengujian yang dilakukan, diperoleh kesimpulan bahwa penggunaan *vector space model* dengan *query expansion* dapat meningkatkan nilai *recall* atau dokumen relevan yang ditampilkan lebih banyak dibandingkan *vector space model* tanpa *query expansion*. Namun, untuk nilai *precision* dan *f-measure* pencarian menggunakan *query expansion* sedikit lebih rendah dibandingkan pencarian tanpa *query expansion*. Pencarian menggunakan *query expansion* memiliki nilai performansi rata-rata *recall* 100%, *precision* 38%, dan *f-measure* 50%, sedangkan pencarian tanpa *query expansion* memiliki nilai performansi rata-rata *recall* 30%, *precision* 19%, dan *f-measure* 16%.

Kata Kunci: Sistem temu kembali informasi, *Vector Space Model*, *Query Expansion*, perangkangan dokumen

ABSTRACT

In web search engines, the number of relevant documents obtained is influenced by the number of keywords in the query. Queries entered by users when performing searches are generally short, consisting of only one or two words. This resulted in only a few relevant documents called in the database due to the lack of information in the query. In this study, the tourism information retrieval system was built using a vector space model and query expansion using the Indonesian Thesaurus dictionary. The vector space model is a model used to store documents in the form of a term-document matrix. Meanwhile, query expansion is a technique to expand the meaning of keywords to meet the information needs needed by the user. Previous web page information is retrieved using a web crawler and stored in a database. Performance measurement is done by calculating the precision, recall and f-measure values from the search results with several different keywords. From the results of the tests conducted, it is concluded that the use of a vector space model with query expansion can increase the recall value or relevant documents that are displayed more than the vector space model without query expansion. However, for precision and f-measure values, the search using query expansion is slightly lower than the search without query expansion. Search using query expansion has an average performance value of 100% recall, 38% precision, and 50% f-measure, while the search without query expansion has an average performance value of 30% recall, 19% precision, and 16% f-measure.

Keywords: *Information Retrieval, Vector Space Model, Query Expansion, ranking documents*